Алгоритм минимизации энергии Гиббса: Расчет химического равновесия*

Ю. В. Заика
Институт прикладных математических исследований КарНЦ РАН, Петрозаводск, Россия
e-mail: zaika@krc.karelia.ru

Предложен вычислительный алгоритм минимизации энергии Гиббса для определения состава смеси в состоянии химического равновесия. Алгоритм применим и для случая многофазных систем. Основное внимание уделено критическим случаям, когда методы оптимизации, основанные на использовании градиента и гессиана, теряют эффективность.

Ключевые слова: численные методы условной оптимизации, свободная энергия Гиббса, химическое равновесие.

1. Постановка задачи

Проблема определения равновесного состава смеси является одной из традиционных в химической термодинамике. Обстоятельный анализ данной темы и библиография к ней содержатся, например, в [1 – 3]. Что касается эффективных численных методов, то, по-видимому, здесь отсчет следует вести с классической работы [4] (см. также статьи [5, 6] и литературные ссылки к ним). Несмотря на наличие развитого программного обеспечения, вычислительные проблемы остаются (универсальный алгоритм невозможен). По оценкам экспертов (автор ориентировался на обзор возможностей пакета HSC Chemistry) примерно для 10 % задач автоматический режим их решения не удовлетворяет возрастающим требованиям и возникает необходимость постоянного совершенствования алгоритмов. Так, например, для идеальной газовой смеси задача определения равновесного состава при фиксированных температуре и давлении состоит в минимизации по переменным n_i энергии Гиббса

$$G = \sum_{i=1}^{k} n_i \big(G_i^0(T) + RT \ln(P n_i \, \bar{n}^{-1})\big) \to \min, \quad \bar{n} \equiv \sum_{i=1}^{k} n_i,$$

при соблюдении линейных ограничений материального баланса. Здесь n_i — количество молей i-й составляющей смеси; G^0 — стандартный химический потенциал; R — универсальная газовая постоянная; T — абсолютная температура; P — давление (число атмосфер). В более общем случае добавится еще одна операция суммирования (по количеству фаз) и под знаком логарифма появятся множители, называемые коэффициентами активности. Не останавливаясь здесь на терминологии и подробностях (за этим

*Работа выполнена при финансовой поддержке РФФИ (грант № 09-01-00439).

следует обратиться к руководствам по физической химии), отметим лишь, что с математической точки зрения при $n_j \to +0$ имеем особенность ($\ln \to -\infty$), что неизбежно сказывается на работоспособности методов оптимизации, основанных на использовании градиента и гессиана. И дело не только в том, что по условиям задачи может потребоваться определение концентраций компонентов менее $10^{-18}-10^{-19}$ [1]. В процессе итераций на допустимом многограннике при большом объеме промежуточных вычислений и количестве переменных, исчисляемом десятками, трудно контролировать влияние на точность решения задачи появлений "почти нулей".

Цель настоящей работы — предложить алгоритм, который учитывает указанную особенность. К элементарным операциям отнесены решение задачи линейного программирования и поиск минимума выпуклой функции на отрезке. Для них имеется множество эффективных алгоритмов (автор пользовался пакетом Scilab 4.1). Чтобы сосредоточиться на основной идее, рассмотрим классическую постановку задачи [2]. Итак, требуется минимизировать выпуклую однородную функцию

$$g(n) = g(n_1, \ldots, n_k) = \sum\nolimits_{i=1}^{k} n_i\big(c_i + \ln(n_i\,\bar{n}^{-1})\big),$$

где c_i — постоянные (равные $G_i^0/RT + \ln P$); n_i — количество молей i-го компонента смеси; $n = (n_1, \ldots, n_k)^\top \in \mathbb{R}^k_* \equiv \{n \neq 0 \,|\, n_i \geqslant 0,\, 1 \leqslant i \leqslant k\}$; $\bar{n} = n_1 + \ldots + n_k = \|n\|$. Верхний индекс $\top$ означает транспонирование, нули линейных пространств обозначаем одним символом, $\|\cdot\|$ — октаэдрическая норма в $\mathbb{R}^k$ (сумма модулей компонент вектора). В дальнейшем считаем $c_i < 0$ (см. пример и замечания), иначе ниже вместо максимальной скорости убывания целевой функции следует говорить о минимальной скорости изменения. Материальный баланс выражается ограничением $A^\top n = b$. Элементы матрицы $A = \{a_{ij}\}_{k\times m}$ — неотрицательные целые числа (множество которых обозначим через $\mathbb{Z}_0$), $\operatorname{rank} A = m$, нулевые строки отсутствуют, $b_j > 0$ $\big(b_j \in \mathbb{R}_+ \equiv \{r > 0\}\big)$, $1 \leqslant j \leqslant m < k$. В скобках после n_i — компоненты градиента g, $(\operatorname{grad} g)_i \to -\infty$ при $n_i \to +0$. В силу ограничений материального баланса сумма молей ограничена как сверху, так и снизу: $0 < \bar{n}_{\min} \leqslant \bar{n} \leqslant \bar{n}_{\max} < +\infty$.

Используя мольные доли $x_i = n_i/\bar{n}$, запишем задачу в следующей форме:

$$g(n) = \bar{n} f(x) \equiv \bar{n}\big(c^\top x + \varphi(x)\big) \to \min, \quad \varphi(x) \equiv \sum_{i=1}^{k} n_i\,\bar{n}^{-1} \ln(n_i\,\bar{n}^{-1}) = \sum_{i=1}^{k} x_i \ln x_i,$$

$$x \in S = \{x \in \mathbb{R}^k |\; x_i \geqslant 0,\; x_1 + \ldots + x_k = 1\}, \quad A^\top n = b, \quad A \in \mathbb{Z}_0^{k\times m}, \quad b \in \mathbb{R}^m_+.$$

В пространстве $\{\mathbb{R}^k, \|\cdot\|\}$ вектор $x = n/\bar{n}$ имеет единичную длину и определяет направление. По непрерывности доопределяем $x_i \ln x_i = 0$ при $x_i = 0$ в силу $\alpha \ln \alpha \to -0$, $\alpha \to +0$. Функцию φ (и f) при необходимости можно считать определенной не только на множестве S: $\varphi(0) = 0$, $\varphi(n) = \varphi(tn) = \varphi(x)$, $t > 0$, $n \in \mathbb{R}^k_*$. Значения $\varphi(n)$ определяются лишь направлением, причем независимо от перестановки компонент n_i вектора n. Итак, целевая функция имеет специальную структуру: сумма молей компонентов смеси умножается на функцию, зависящую лишь от распределения мольных долей.

2. Грубая оценка минимума и направления спуска

Экстремумы функции $\varphi(x)$ на S находятся аналитически: $\max \varphi = 0$, $\min \varphi = -\ln k$. Максимум достигается на базисных векторах $e^i = (0, \ldots, 1, \ldots, 0)^\top$ (смесь вырождается

в компонент). Минимум единственный и достигается на равномерном распределении мольных долей $x_i = 1/k$. Итак, целевая функция имеет двустороннюю оценку:

$$L_0(n) \leqslant g(n) \leqslant L^0(n), \quad L_0(n) \equiv c^\top n - \bar{n}\ln k,\ L^0(n) \equiv c^\top n.$$

Допустимое множество $D = \{n \,|\, n_i \geqslant 0,\ A^\top n = b\}$ компактно (выпуклый многогранник), минимальное значение $g_* = \min g$ оценивается решением двух задач линейного программирования: $g_* \in [g_*^-, g_*^+]$, $g_*^- \equiv \min L_0$, $g_*^+ \equiv \min L^0$, $n \in D$. Вектор целевой функции L_0 отличается от вектора c равномерной поправкой компонент на $-\ln k$.

Рассмотрим задачу $f(x) = c^\top x + \varphi(x) \to \min$, $x \in S$. Поясним ее смысл. Пусть формально сумма $\bar{n}$ фиксирована и ограничение $A^\top n = b$ не принимается в расчет. Тогда оптимальное распределение долей x_i определяется именно задачей $f \to \min$. Приведем аргумент в геометрических терминах. Фиксируем единичный направляющий вектор: $n^0 \in S$ ($\bar{n}^0 = 1$). На луче $n(t) = tn^0$ ($t \geqslant 0$ интерпретируем как время движения) имеем $g(n(t)) = t\big(c^\top n^0 + \varphi(n^0)\big)$. Производная по t равна $f(n^0)$. Целесообразно выбрать направление n^0, вдоль которого функция g убывает ($f < 0$) наискорейшим образом: $f(n^0) \to \min$, $n^0 \in S$. Величина $g(n)$ определяется как значение f на направлении $n^0 = n/\bar{n}$ (распределение мольных долей), умноженное на время $t = \bar{n}$ движения из нуля вдоль n^0 в точку n. Ограничение $A^\top n = b$ определяет компромисс между стремлением к быстрому убыванию g и увеличением времени встречи с множеством D.

Фиксируем номер наименьшего c_i. В общем случае это номер одного из наименьших c_i, который без ограничения общности считаем равным k. Для упрощения изложения далее подобные оговорки опускаем. Выразив в $f(x)$ переменную $x_k = 1 - x_1 - \ldots - x_{k-1}$, получим (с учетом $\alpha = 0 \Rightarrow \alpha\ln\alpha = 0$) функцию $F(y)$, $y = (x_1, \ldots, x_{k-1})^\top \in [0,1]^{k-1}$. На множестве $(0,1)^{k-1}$ функция F строго выпукла, поскольку гессиан положительно определен: $F''_{x_ix_i} = 1/x_i + 1/x_k$, $F''_{x_ix_j} = 1/x_k$, $1 \leqslant i, j \leqslant k-1$, $i \neq j$. Стационарная точка в $(0,1)^{k-1}$ будет единственным минимумом F на $[0,1]^{k-1}$. Приравнивая производные функции F по x_i к нулю, приходим к системе линейных уравнений

$$\big(1+\exp(c_1-c_k)\big)x_1 + x_2 + \ldots + x_{k-1} = 1,\ \ldots,\ x_1 + \ldots + x_{k-2} + \big(1+\exp(c_{k-1}-c_k)\big)x_{k-1} = 1.$$

Вычитаем первое уравнение из остальных. Затем последовательно "идем снизу вверх": $x_{k-1} = x_1\exp(c_1 - c_{k-1})$, $\ldots$, $x_2 = x_1\exp(c_1 - c_2)$. Подставляя полученные выражения в первое уравнение, имеем решение

$$x_i^0 = \big(\exp(c_i - c_1) + \ldots + \exp(c_i - c_k)\big)^{-1}, \quad 1 \leqslant i \leqslant k-1.$$

Затем вычисляем $x_k^0 = 1 - x_1^0 - \ldots - x_{k-1}^0$, $x^0 \in (0,1)^k$. Значение x_k^0 получается и подстановкой $i = k$. С ростом доминирования k-го компонента в смысле $c_k \ll c_i < 0$ $\big(\exp(c_i - c_k) \gg 1\ \forall i \neq k\big)$ имеем $x_i^0 \approx 1/\exp(c_i - c_k)$ и в пределе получаем $x_i = 0$, $1 \leqslant i \leqslant k-1$, $x_k = 1$, $f = c_k$. Смесь содержит практически один (сильно доминирующий) компонент. Если, например, имеются два сильно доминирующие компонента $\big(c_k = c_s \ll c_i < 0,\ i \notin \{k, s\}\big)$, то $x_k^0 = x_s^0 \approx 1/2$, $f \approx c_k$. Другая крайность — отсутствие доминирования: $c_1 = \ldots = c_k \equiv \beta$. Тогда $x_i^0 = 1/k$. При этом $g(n) = \bar{n}\big(\beta + \varphi(x)\big)$, минимум φ реализуется на равномерном распределении. Найденная точка $x^0 \in (0,1)^k$ строгого минимума f на S определяет направление n^0 наискорейшего убывания $g(n)$.

Теперь примем во внимание ограничение $A^\top n = b$. Двигаясь по экстремальному лучу tn^0 ($t > 0$, $x_i^0 > 0$), пройдем в общем случае мимо допустимого множества D.

Поэтому рассмотрим другие варианты направлений движения. Пусть n^1, n^2 — решения задач линейного программирования (ЛП) $\bar{n} \to \min$, $\bar{n} \to \max$, $n \in D$. Следовательно, известен диапазон значений общего количества молей $\bar{n} = \|n\|$. Для решения задачи n_* имеем оценку $\bar{n}_* \in [\bar{n}_{\min}, \bar{n}_{\max}]$, $\bar{n}_{\min} \equiv \bar{n}^1$, $\bar{n}_{\max} \equiv \bar{n}^2$, $0 < \bar{n}^1 \leqslant \bar{n}^2 < +\infty$. Двигаясь вдоль направлений, в точку n^1 попадаем за минимальное время $t = \bar{n}$, в точку n^2 — за максимальное. Обратимся к структуре целевой функции: $g(n) = \bar{n} f(x)$. Пусть формально x — фиксированный векторный параметр, $f(x) < 0$. Тогда задача $g \to \min$ эквивалентна $\bar{n} \to \max$, $n \in D$. С другой стороны, в силу $A^\top x = b/\bar{n}$ уменьшение $\bar{n}$ ведет к росту компонент вектора x. Это может оказаться предпочтительнее, поскольку $f(x) = c^\top x + \varphi$, $c_i < 0$ (когда $|c_i|$ достаточно велики). Определим точку $\tilde{n}$ минимума $g(n)$ на отрезке $[n^1, n^2] = \lambda n^2 + (1-\lambda) n^1$, $\lambda \in [0,1]$, и предварительно фиксируем направление $\tilde{n}^0 = \tilde{x}^0 = \tilde{n}/\|\tilde{n}\|$. Помимо n^1, n^2 следует рассмотреть точки $n^3, \ldots, n^6 \in D$ экстремумов оценочных функций $L_0(n)$, $L^0(n)$. Поскольку $\varphi(n_*) \in [-\ln k, 0\,]$, то можно использовать и "промежуточные" функции $c^\top n - \lambda_s \bar{n} \ln k$, $n \in D$, $\lambda_s \in (0,1)$. Максимумы рассматриваем в силу того, что векторы из максимума в минимум в линейном приближении могут претендовать на направления спуска. Минимум из минимумов g на невырожденных отрезках $[n^i, n^j]$ определяет направление $\tilde{n}^0 = d^0$ (не обязательно единственное).

Итак, имеем векторы n^0, d^0 $\big(f(n^0) \leqslant f(d^0) < 0\big)$. На луче tn^0 целевая функция убывает максимально без учета $A^\top n = b$. Второе направление позволяет попасть в D (при соответствующем значении $t = \|\tilde{n}\|$), но в общем случае с меньшей (по абсолютной величине) скоростью убывания. Можно выбирать и компромиссные направления: $h = \lambda n^0 + (1-\lambda) d^0$, $\lambda \in [0,1]$. Логично также добавить равномерное направление $u^0 = (1/k, \ldots, 1/k)^\top$, которое минимизирует функцию $\varphi(x)$: $h = \lambda_1 n^0 + \lambda_2 d^0 + \lambda_3 u^0$, где $\lambda_i \geqslant 0$, $\lambda_1 + \lambda_2 + \lambda_3 = 1$. У векторов n^0, d^0 могут быть нулевые компоненты с равными номерами. Тогда *a priori* игнорируются некоторые составляющие смеси. Чтобы $h \in \mathbb{R}^k_+$ $(h_i > 0)$, достаточно брать $\lambda_3 > 0$, принимая u^0 как стабилизатор: $0 < \lambda_3 \ll 1$.

3. Первое приближение

Фиксируем направление $h \in S \cap \mathbb{R}^k_+$. Целесообразно начинать с предпочтения d^0 $(\lambda_2 \approx 1)$ в выпуклой комбинации n^0, d^0, u^0. Двигаемся по лучу th, $t \geqslant 0$, t интерпретируем как время. Обозначим через $a^j \in \mathbb{Z}^k_0$ столбцы матрицы A: $\langle a^j, n \rangle = b_j$, $1 \leqslant j \leqslant m$. Угловые скобки означают скалярное произведение $\langle x, y \rangle = x_1 y_1 + \ldots + x_k y_k$, $|\cdot| = \langle \cdot, \cdot \rangle^{1/2}$ — длина. Если подставить $n = th$ в ограничения, то из $t\langle a^j, h \rangle = b_j$ определяются моменты времени $t_j = b_j \langle a^j, h \rangle^{-1} > 0$, $1 \leqslant j \leqslant m$, при которых достигается материальный баланс по каждому элементу. Сначала воспользуемся методом наименьших квадратов:

$$\rho^2 \equiv m^{-1} \sum_{j=1}^{m} \big(t\langle a^j, h \rangle - b_j\big)^2 \to \min \;\Rightarrow\; t_0 \equiv t_{\min} = \langle Ab, h \rangle \big| A^\top h \big|^{-2} > 0.$$

Значения c_j, b_j заданы приближенно, поэтому $n = t_0 h$ может оказаться приемлемым решением задачи при достаточно малом среднеквадратичном уклонении ρ. При необходимости жесткого соблюдения ограничений вектор n спроектируем ортогонально на линейное многообразие $\Lambda = \{z \in \mathbb{R}^k |\, A^\top z = b\}$:

$$p = n - A(A^\top A)^{-1}(A^\top n - b) = t_0 H + B, \quad H \equiv h - A(A^\top A)^{-1} A^\top h,\; B \equiv A(A^\top A)^{-1} b.$$

Здесь H — ортогональная проекция вектора h на $\Lambda_0 = \{z \in \mathbb{R}^k | \, A^\top z = 0\}$, B — перпендикуляр (см., например, [7]), являющийся решением системы $A^\top z = b$ с минимальной длиной. Если $p \in \mathbb{R}^k_*$, то $p \in D$ и можно за первое приближение принять $n^1_* = p$. В общем случае, не исключая появления в результате вычислений отрицательных компонент $p_j < 0$, дополнительно проектируем на неотрицательный ортант (переопределяем $p_j := 0$). Получаем вектор p^0, хотя при этом, строго говоря, $n^1_* = p^0 \notin D$. Проектирование последовательно на многообразие Λ и ортант следует повторить несколько раз.

При более детальном рассмотрении действуем следующим образом. Проектируем движущуюся точку th, $t \geqslant 0$, ортогонально на линейное многообразие Λ: $p(t) = tH + B$. Для уточнения значений t учтем свойства матрицы материального баланса A. Сумма столбцов a^j есть вектор $a^+ > 0$ с положительными компонентами, поэтому у вектора H есть отрицательные компоненты H_i: $\langle a^+, H\rangle = 0$. В силу $A^\top B = b$ имеются положительные компоненты B_j, в невырожденном случае $B > 0$. Векторное ограничение $p(t) \geqslant 0$ (неравенства вида $\alpha_i t \geqslant \beta_i$) даст отрезок $[t_1, t_2]$, на котором $p(t) \in D$. В качестве первого приближения n^1_* берем точку минимума $g(p(t))$, $t \in [t_1, t_2]$. Отметим, что формально не исключено, что $\{t\}$ пусто. Но на текущем этапе алгоритма нет необходимости в строгом включении $p(t) \in D$ (тем более, что вычисления приближенные). В самом неблагоприятном случае останавливаемся на выборе $n^1_* = p^0$.

4. Итерационный процесс

Вначале остановимся на раскрытии неопределенности $\alpha \ln \alpha$ ($\alpha \to +0$) численно. В оригинале [4] приводятся шесть знаков после запятой, т.е. точнее задачу решать не требовалось. Условно считаем, что (в конкретной задаче) мольные доли $\alpha = x_i$ менее 10^{-7} представляют лишь теоретический интерес. Фиксируем соответствующее исчезающе малое по смыслу задачи $\varepsilon > 0$ (пусть $\varepsilon{=}10^{-10}$). В вычислениях полагаем $\alpha \in (-\varepsilon, \varepsilon) \;\Rightarrow\; \alpha \ln \alpha = 0$. При необходимости можно взять отрезок ряда для функции $\alpha \ln \alpha$, $\alpha \in [0, \varepsilon)$.

Следующий шаг — анализ ситуации $(\operatorname{grad} g(n))_i = c_i + \ln x_i \to -\infty$, $x_i \to +0$. Образно говоря, алгоритму градиентного типа хочется шагнуть в этом направлении, но чтобы удержаться в рамках ограничений, приходится умножать большие числа на малые с потерей точности вычислений и непредсказуемыми последствиями для сходимости (в практически важных задачах число переменных доходит до десятков). Если бы выпуклая функция $g(n)$ допускала дифференцируемое продолжение в окрестность множества D, то можно было бы воспользоваться известным критерием оптимальности: $\min\langle \operatorname{grad} g(n_*), n\rangle = \langle \operatorname{grad} g(n_*), n_*\rangle$, $n \in D$ [7]. Для текущего s-го приближения $n^s_* \approx n_*$ находится решение $n = \tilde{n}^s_*$ задачи линейного программирования $\langle \operatorname{grad} g(n^s_*), n\rangle \to \min$, $n \in D$. Если в точке n^s_* критерий оптимальности не выполняется, то вектор $\tilde{n}^s_* - n^s_*$ указывает направление строгого убывания целевой функции g и приближение n^{s+1}_* определяется минимумом g на отрезке $[n^s_*, \tilde{n}^s_*]$, причем $g(n^{s+1}_*) < g(n^s_*)$.

В рассматриваемой задаче неопределенность в правой части критерия оптимальности раскрывается в силу однородности целевой функции g: $\langle \operatorname{grad} g(n), n\rangle = g(n)$. Ориентируясь на локальную аппроксимацию $g(n) = c^\top n + \bar{n}\varphi(n) \approx c^\top n + \bar{n}\varphi(n^s_*)$ ($\varphi(0) = 0$, $\varphi(n) = \varphi(n/\bar{n}) = \varphi(x)$), рассмотрим задачу линейного программирования

$$L^-_s(n) = c^\top n + \bar{n}\varphi(n^s_*) = \sum_{i=1}^{k} \big(c_i + \varphi(n^s_*)\big) n_i \to \min, \quad n \in D.$$

По сравнению с вариантом $L_s^+(n) = \langle \operatorname{grad} g(n_*^s), n\rangle = \sum\limits_{i=1}^{k}(c_i + \ln x_{*i}^s)n_i \to \min$ диапазон изменения коэффициентов линейной формы сужается с несобственного множества $[-\infty, c_i]$ до равномерной поправки $\varphi(n_*^s) \in [-\ln k, 0]$. Для хорошего первого приближения n_*^1 итерации на основе линейной аппроксимации L_s^- (последовательный переход $n_*^s \rightarrowtail \tilde{n}_*^s \rightarrowtail n_*^{s+1}$) могут привести к решению задачи.

Попытаемся расширить возможности аппроксимации, объединив невырожденность формы L_s^- с экстремальными свойствами L_s^+. Ограничим множество возможных значений коэффициентов линейной формы, поставив барьер неограниченному росту нормы градиента. Для этого нужен масштаб скорости. Логично воспользоваться оценкой $L_0(n) \leqslant g(n) \leqslant L^0(n)$, $\operatorname{grad} L_0 = c - \ln k$ (покомпонентно $c_i - \ln k$), $\operatorname{grad} L^0 = c$, фиксировать максимальную по абсолютной величине скорость убывания $V \equiv c_k - \ln k$ ($c_k = \min c_i < 0$) и не позволять недоминирующим компонентам двигаться существенно быстрее (локально, в линейном приближении). Чем меньше зазор между гиперплоскостями $\big(\max(L^0 - L_0) = \bar{n}_{\max} \ln k,\ n \in D\big)$, тем обоснованнее такое ограничение скорости.

Обратим внимание на следующее обстоятельство. При классической линеаризации (форма L_s^+) в слагаемых целевой функции $(c_i + \ln x_i)n_i$ текущее приближение используется для фиксации нелинейности, зависящей явно лишь от мольной доли (замена переменной x_i на значение x_{*i}^s). Будем придерживаться этой схемы с той лишь разницей, что вследствие возможности $\ln x_i \to -\infty$ сначала выделим функции $x_i \ln x_i$ (разрешимую особенность), оставляя n_i свободными переменными после подстановки x_{*i}^s вместо x_i. Формализуем приведенные соображения.

Определим вектор $\tilde{c}$, состоящий из коэффициентов линейной аппроксимирующей формы $L_s(n) = \tilde{c}^\top n$, алгоритмически. Предварительно обнулим массив: $\tilde{c} = 0 \in \mathbb{R}^k$. Если $x_*^s = n_*^s/\bar{n}_*^s = e^i$ (вырожденное приближение смеси одной составляющей), то полагаем $\tilde{c} = c$, поскольку локально $g(n) = \bar{n}\big(c^\top x + \varphi(x)\big) \approx \bar{n}\big(c^\top x + \varphi(x_*^s)\big) = c^\top n$. При этом $L_s(n) = c^\top n = L_s^-(n) = L^0(n)$ — верхняя оценка $g(n)$. Классическая форма L_s^+ не имеет смысла ($\ln = -\infty$). Далее уже считаем, что среди мольных долей x_{*i}^s нет единицы.

Шаг 1. Рассмотрим первое слагаемое в функции $g(n)$, явно выделив разрешимую особенность: $n_1(c_1 + \ln x_1) = n_1 c_1 + \bar{n} x_1 \ln x_1$, $x_1 = n_1/\bar{n}$, $\bar{n} = n_1 + \ldots + n_k$. Если в каждом слагаемом $g(n)$ заменить x_i на x_{*i}^s, то в сумме получим линейную форму $L_s^-(n)$ с достаточно узким диапазоном значений коэффициентов (отрезки $[c_i - \ln k, c_i]$). Задача состоит в расширении этого диапазона, но не до несобственных множеств $[-\infty, c_i]$, как в аппроксимации $L_s^+(n)$: $L_s^+(n) = \langle \operatorname{grad} g(n_*^s), n\rangle$, $(\operatorname{grad} g)_i = c_i + \ln x_{*i}^s \in [-\infty, c_i]$. Если $x_{*1}^s = 0\,(< \varepsilon)$, то в правой части тождества $n_1(c_1 + \ln x_1) = n_1 c_1 + \bar{n} x_1 \ln x_1$ осуществляем подстановку значения x_{*1}^s текущего приближения в особенность $x_1 \ln x_1$ вместо переменной x_1. В силу $x_{*1}^s \ln x_{*1}^s = 0$ в форме $L_s(n) = \tilde{c}^\top n$ коэффициент при n_1 полагаем равным $\tilde{c}_1 = c_1$. Пусть $x_{*1}^s > 0\,(> \varepsilon)$. Если выполняется неравенство

$$\xi_1 \equiv V(c_1 + \ln x_{*1}^s)^{-1} \geqslant 1 \ \Leftrightarrow\ (\operatorname{grad} g(n_*^s))_1 \geqslant V, \quad V \equiv c_k - \ln k < 0,$$

то присваиваем $\tilde{c}_1 := c_1 + \ln x_{*1}^s$ (как и в L_s^+). Выполнение неравенства означает, что первая компонента градиента по абсолютной величине не превышает порога скорости $|V|$, определенного потенциально доминирующим компонентом смеси. Остается рассмотреть случай $\xi_1 \in (0, 1)$. При $\xi_1 = +0$ ($x_{*1}^s = +0$) и $\xi_1 = 1$ значения коэффициента $\tilde{c}_1$ уже определены: это c_1 и V. Поэтому естественно потребовать выполнения предельных переходов $x_{*1}^s \to +0 \Rightarrow \tilde{c}_1 \to c_1$, $\xi_1 \to 1 - 0 \Rightarrow \tilde{c}_1 \to V$.

Проведем формальные преобразования. В тождество

$$n_1(c_1+\ln x_1)=\varsigma n_1(c_1+\ln x_1)+(1-\varsigma)(n_1c_1+\bar{n}x_1\ln x_1)$$

подставим значения $\varsigma=\xi_1^2$, $x_1=x_{*1}^s$. В правой части получаем выражение $\xi_1 V n_1+(1-\xi_1^2)(n_1c_1+\bar{n}x_{*1}^s\ln x_{*1}^s)$. Параметр ς выбран именно для согласования при указанных предельных переходах (для этих целей допустимы и $\varsigma=\xi_1^{1+\nu_1}$, $\nu_1>0$). При переменной n_1 фиксируем множитель $\tilde{c}_1=\xi_1 V+(1-\xi_1^2)(c_1+x_{*1}^s\ln x_{*1}^s)$. Требование выполнено. Дополнительно из-за множителя $\bar{n}=n_1+\ldots+n_k$ следует для $j=2,\ldots,k$ переопределить значения $\tilde{c}_j:=\tilde{c}_j+(1-\varsigma)x_{*1}^s\ln x_{*1}^s=(1-\varsigma)x_{*1}^s\ln x_{*1}^s$ (до шага 1 $\tilde{c}_j=0$).

Шаг 2. Переходим к слагаемому $n_2(c_2+\ln x_2)=n_2c_2+\bar{n}x_2\ln x_2$. Если $x_{*2}^s=0$, то в силу $x_{*2}^s\ln x_{*2}^s=0$ к определенному на шаге 1 значению $\tilde{c}_2$ добавляем величину c_2: $\tilde{c}_2:=\tilde{c}_2+c_2$. Пусть $x_{*2}^s>0$. Если $\xi_2\equiv V/(c_2+\ln x_{*2}^s)\geqslant 1$ (скорость убывания медленнее принятого порога), то к $\tilde{c}_2$ (определенном на шаге 1) добавляем число $c_2+\ln x_{*2}^s$ в соответствии с левой частью равенства. Остается рассмотреть случай $\xi_2\in(0,1)$. В тождество

$$n_2(c_2+\ln x_2)=\varsigma n_2(c_2+\ln x_2)+(1-\varsigma)(n_2c_2+\bar{n}x_2\ln x_2)$$

подставляем значения $\varsigma=\xi_2^2$, $x_2=x_{*2}^s$: $\xi_2 V n_2+(1-\xi_2^2)(n_2c_2+\bar{n}x_{*2}^s\ln x_{*2}^s)$. К значению $\tilde{c}_2$ добавляем $\Delta\tilde{c}_2=\xi_2 V+(1-\xi_2^2)(c_2+x_{*2}^s\ln x_{*2}^s)$. Кроме того, к $\tilde{c}_1$ и $\tilde{c}_j$ $(j=3,\ldots,k)$ добавляем число $(1-\xi_2^2)x_{*2}^s\ln x_{*2}^s$. Отметим, что если на шаге 1 $\tilde{c}_1=c_1$ (как в верхней оценке $L^0(n)$), то отрицательная добавка $(1-\xi_2^2)x_{*2}^s\ln x_{*2}^s$ "в нужном направлении". Аналогичны выкладки для варианта $\varsigma=\xi_2^{1+\nu_2}$, $\nu_2>0$.

Продолжая процесс преобразования слагаемых $n_i(c_i+\ln x_i)=n_ic_i+\bar{n}x_i\ln x_i$ последовательно, сформируем вектор $\tilde{c}$ и определим аппроксимирующую форму $L_s(n)=\tilde{c}^\top n$. Заметим, что помимо $L_s=L^0$ (когда найдется $x_{*i}^s=1$) реализуется и нижняя оценка функции g: $x_*^s=(1/k,\ldots,1/k)^\top\ \Rightarrow\ L_s=L_s^-=L_0$. Когда все $\xi_i\geqslant 1$ (принятый ориентир $|V|$ скорости убывания не превышен), то $L_s=L_s^+=\langle\operatorname{grad} g(n_*^s),n\rangle$. Неограниченный рост $|\tilde{c}_i|$ исключен, поскольку особенность $\alpha\ln\alpha$ выделена явно и ограничена. Формирование вектора $\tilde{c}$ обобщенно можно считать регуляризацией $\operatorname{grad} g$. Цель расширения множества линейных форм $\big(\{L_s^-\}\subset\{L_s\}\subset\{L_s^+\}\big)$ достигнута (но предложенный формализм преобразований с $\alpha\ln\alpha$ не единственно возможный).

Далее переход от текущего приближения $n_*^s\approx n_*$ к следующему стандартен. Решение $\tilde{n}_*^{s+1}$ задачи $L_s(n)=\tilde{c}^\top n\to\min$, $n\in D$, соединяем отрезком с вектором n_*^s. Приближение n_*^{s+1} определяется минимумом функции $g(\lambda\tilde{n}_*^{s+1}+(1-\lambda)n_*^s)$, $\lambda\in[0,1]$. Изложим некоторую модификацию: с учетом "кривизны" траектории $\{x_*^i\}$ целесообразно дополнительно рассмотреть отрезки, соединяющие $\tilde{n}_*^{s+1}$ с $\arg\max L_s$ и, например, с n_*^{s-1}. Остановимся на следующем варианте. Находим минимумы m^1,m^2 целевой функции $g(n)$ на отрезках $[\arg\min L_s,\arg\max L_s]$, $[\arg\min L_s,n_*^s]$. Следующее приближение n_*^{s+1} выбираем как $\arg\min g(n)$, $n\in[m^1,m^2]$. Критерием остановки может служить $\min L_s(n)=L_s(n_*^s)$, $n\in D$, или достижение заданной относительной погрешности: $\big(g(n_*^{s+1})-g(n_*^s)\big)/g_*^{+,-}<\delta$, $\big(g(n_*^{s+1})-g(n_*^s)\big)/(g_*^--g_*^+)<\delta$.

Отметим, что если на некоторой итерации $n_*^{s+1}=m^1\in[\arg\min L_s,\arg\max L_s]\subset D$, то последующие приближения принадлежат D (без учета погрешностей вычислений). В любом случае двигаемся из n_*^s в направлении D, поэтому начального включения $n_*^1\in D$ нет необходимости придерживаться слишком строго. Если при $n_*^s\notin D$ оказалось $n_*^{s+1}=n_*^s$, то сдвигаем n_*^s внутрь отрезка $[\arg\min L_s,n_*^s]$ (в сторону D, $\arg\min L_s\in D$) и повторяем цикл либо переходим в допустимое множество D: $n_*^{s+1}=m^1$.

Квадратичное приближение. В случае большой размерности задачи можно наращивать влияние нелинейности методом продолжения по параметру:

$$g(n,\tau) \equiv \sum_{i=1}^{k} \left(c_i n_i + \tau n_i \ln(n_i \bar{n}^{-1})\right) \to \min, \quad A^\top n = b, \quad n \in \mathbb{R}^k_*, \quad \tau = \tau_1 > 0, \ldots, \tau = 1.$$

Целесообразна также следующая декомпозиция. Заметим, что $n \leftrightarrow \{\bar{n}, x\}$. Это позволяет перейти к формально скалярной экстремальной задаче $\bar{n}\psi(\bar{n}) \to \min$, $\bar{n} \in [\bar{n}_{\min}, \bar{n}_{\max}]$. Здесь значения функции $\psi(\bar{n})$ заданы алгоритмически как решения выпуклой задачи сепарабельного программирования с параметром $\bar{n}$:

$$f(x) = c^\top x + \varphi(x) \to \min, \quad A^\top x = b\bar{n}^{-1}, \quad x_1 + \ldots + x_k = 1, \quad x_i \geqslant 0.$$

График функции $y(\alpha) = \alpha \ln \alpha$ похож на параболу. В силу этого в зависимости от приближения n^s_* слагаемые $x_i \ln x_i$ в функции $\varphi(x)$ аппроксимируются параболами: если компонента вектора n^s_* не слишком близка к граничным точкам отрезка $[0,1]$, то парабола строится по трем точкам, иначе используем информацию $y(1/e) = -1/e$, $y'(1/e) = 0$. Промежуточное приближение $\tilde{n}^{s+1}_*$ определяется приближенным решением задачи сепарабельного квадратичного программирования с последующими итерациями проектирования на линейное многообразие Λ и неотрицательный ортант.

Многофазная задача. В этом случае принципиальных изменений схема вычислений не претерпевает. Целевая функция аддитивна: $n = (n^{(1)\top}, \ldots, n^{(r)\top})^\top$, $g(n) = g^{(1)}(n^{(1)}) + \ldots + g^{(r)}(n^{(r)})$. Если j-я фаза однокомпонентна, то формально полагаем $x^{(j)} = 1$, $h^{(j)} = 1$, скалярная переменная $n^{(j)}$ входит в общую целевую функцию $g(n)$ линейно. Однако подчеркнем, что при формировании распределения мольных долей компонентов внутри каждой фазы будут участвовать все компоненты вектора n. Здесь основная проблема — размерность задачи.

Пусть имеется приближенное решение n многофазной задачи, полученное путем применения представленного алгоритма на основе корректировки линейных аппроксимаций. Основной целью этого алгоритма теперь считаем определение общей суммы молей смеси $\bar{n} = \bar{n}^{(1)} + \ldots + \bar{n}^{(r)}$. Векторная составляющая $n^{(i)}$ текущего приближения позволяет "вырезать" из ограничения $A^\top n = b$ соответствующую часть $A^{\top(i)} n^{(i)} = b^{(i)}$. Если среди компонент $b^{(i)}_j$ вектора $b^{(i)}$ есть нулевые, то понижаем размерность подзадачи, удаляя j-ю строку и столбцы с ненулевыми $A^{\top(i)}_{js}$. Делим на $\bar{n}^{(i)}$ (фиксированное число) и "раскрепощаем" мольные доли. Далее перераспределяем мольные доли в рамках каждой многокомпонентной фазы в соответствии с сепарабельным квадратичным алгоритмом. Получаем новое приближение n (комбинированный итерационный алгоритм линейной аппроксимации системы в целом и квадратичной аппроксимации "внутри" многокомпонентных фаз).

5. Пример

В качестве примера рассмотрим задачу из [4], решение которой другими методами известно, что позволяет сравнить результаты. Данные представлены в таблице.

Третья составляющая смеси (H_2O) является сильно доминирующей: направление $n^0 = x^0$ ($x^0_1 + \ldots + x^0_k = 1$) наискорейшего убывания целевой функции $g(n)$ без учета ограничений $A^\top n = b$ имеет третью компоненту 0.999 (ограничимся тремя цифрами

Исходные данные задачи, $\ln P = 3.932$

i	Компонент	G^0/RT	H, a_{i1}	N, a_{i2}	O, a_{i3}
1	H	-10.021	1	0	0
2	H_2	-21.096	2	0	0
3	H_2O	-37.986	2	0	1
4	N	-9.846	0	1	0
5	N_2	-28.653	0	2	0
6	NH	-18.918	1	1	0
7	NO	-28.032	0	1	1
8	O	-14.640	0	0	1
9	O_2	-30.594	0	0	2
10	OH	-26.111	1	0	1
	b_j	—	2	1	1

после точки). Минимумы линейных функций $\bar{n}, L_0, L^0$ достигаются в n^1 (третья компонента 1, пятая 0.5, нули не упоминаем), максимумы — в n^2 (первая компонента 2, четвертая 1, восьмая 1). Получаем предварительные грубые оценки: $\bar{n}_* \in [\bar{n}_{\min}, \bar{n}_{\max}] = [1.5, 4]$, $g_* \in [g_*^-, g_*^+] = [-49.868, -46.414]$. Минимум $g(n)$ на отрезке $[n^1, n^2]$ равен -47.41 и достигается на векторе (0.002, 0, 0.989, 0.01, 0.495, 0, 0, 0.01, 0, 0), нормируя который, получаем d^0. Выберем направление спуска $h = \lambda_1 n^0 + \lambda_2 d^0 + \lambda_3 u^0$, отдавая предпочтение d^0 (например, $\lambda_1 = \lambda_3 = 1/30$). Метод наименьших квадратов дает $t_0 = 1.53$, причем среднеквадратичная невязка достаточно мала: $\rho = 0.025$. Проекция $t_0 h$ на линейное многообразие $\Lambda = \{z \in R^k \,|\, A^\top z = b\}$ равна $p = (0.02,\ 0.004,\ 0.978,\ 0.02,\ 0.48,\ 0.01,\ 0.008,\ 0.01,\ 0,\ 0.002)$ и $g(p) = -47.434$. Уточнение (на отрезке $[t_1, t_2]$) приводит к первому приближению $n_*^1 = (0.03,\ 0.026,\ 0.935,\ 0.027,\ 0.462,\ 0.029,\ 0.02,\ 0.015,\ 0.007,\ 0.017)$, $g(n_*^1) = -47.466$. Далее реализуется итерационный процесс. Первая итерация дает уточнение $g_* \approx -47.66$. Получаем установившийся результат с точностью до тысячных: $n_* \approx (0.04,\ 0.148,\ 0.783,\ 0.0015,\ 0.485,\ 0.002,\ 0.028,\ 0.018,\ 0.0375,\ 0.097)$, $g_* \approx -47.76$.

6. Замечания

1. Задачи малой размерности достаточно хорошо изучены и допускают наглядную интерпретацию в форме графиков и диаграмм. Изложенный алгоритм нацелен на задачи большой размерности, причем на тот встречающийся на практике случай, когда особенность задачи из-за наличия малых мольных долей (в том числе и в промежуточных подзадачах) становится существенной при большом объеме приближенных вычислений. При этом с учетом однородности целевой функции и линейных ограничений следует предварительно нормировать задачу (n заменить на νn) с тем, чтобы, поделив на $\nu = \min b_i$, прийти к такому диапазону молей химических элементов ($\min b_i = 1$), при котором легче понять, что следует принять за "нуль". Конечно, точность вычислений и результатов следует согласовать с точностью входных данных, но этот вопрос в математической постановке представляется труднообозримым.

2. Пусть существует $c_j > 0$. Тогда представим целевую функцию в следующей форме:

$$g(n) = \bar{n}\beta + g_0(n), \quad g_0(n) \equiv \bar{n} f_0(x), \quad f_0(x) \equiv c_0^\top x + \varphi(x), \quad c_{0i} < 0, \quad \beta > \max c_j.$$

Далее ищем направление наискорейшего убывания g_0, как описано выше.

3. Помимо направлений спуска h, регулируемыми параметрами являются порог скорости убывания V (или даже пороги V_i для каждого компонента) и показатели $\nu_i > 0$ скоростей согласования предельных переходов ($\varsigma = \xi_i^{1+\nu_i}$) при построении аппроксимирующей линейной формы $L_s = \tilde{c}^\top n$.

4. Представленный алгоритм не претендует на высокую скорость сходимости. Метод множителей Лагранжа не используется (возможно, это не недостаток). Предложенную схему вычислений можно реализовать как блок генерации начального приближения в алгоритмах более высокого порядка (обычно требующих хорошего начального приближения с положительными компонентами) с целью избежать вырождения вдали от минимума.

Список литературы

[1] Карпов И.К. Физико-химическое моделирование на ЭВМ в геохимии. Новосибирск: Наука, 1981.

[2] Уэйлес С. Фазовые равновесия в химической технологии: Пер. с англ. Ч. 1, 2. М.: Мир, 1989.

[3] Thermodynamic Equilibria and Extrema / A.N. Gorban, B.M. Kaganovich, S.P. Filippov, A.V. Keiko, V.A. Shamansky, I.A. Shirkalin. Springer, 2006.

[4] White W.B., Johnson S.M., Dantzig G.B. Chemical equilibrium in complex mixtures // J. Chem. Phys. 1958. Vol. 28, No. 5. P. 751–755.

[5] Weber C.F. Convergence of the equilibrium code SOLGASMIX // J. Comput. Phys. 1998. Vol. 145. P. 655–670.

[6] Davies R.H., Dinsdale A.T., Gisby J.A. et al. MTDATA — thermodynamic and phase equilibrium software from the National Physical Laboratory // CALPHAD. 2002. Vol. 26, No. 2. P. 229–271.

[7] Аттетков А.В., Галкин С.В., Зарубин В.С. Методы оптимизации: Учеб. для вузов. М.: Изд-во МГТУ, 2003.

Поступила в редакцию 28 января 2010 г.,
с доработки — 20 сентября 2010 г.